

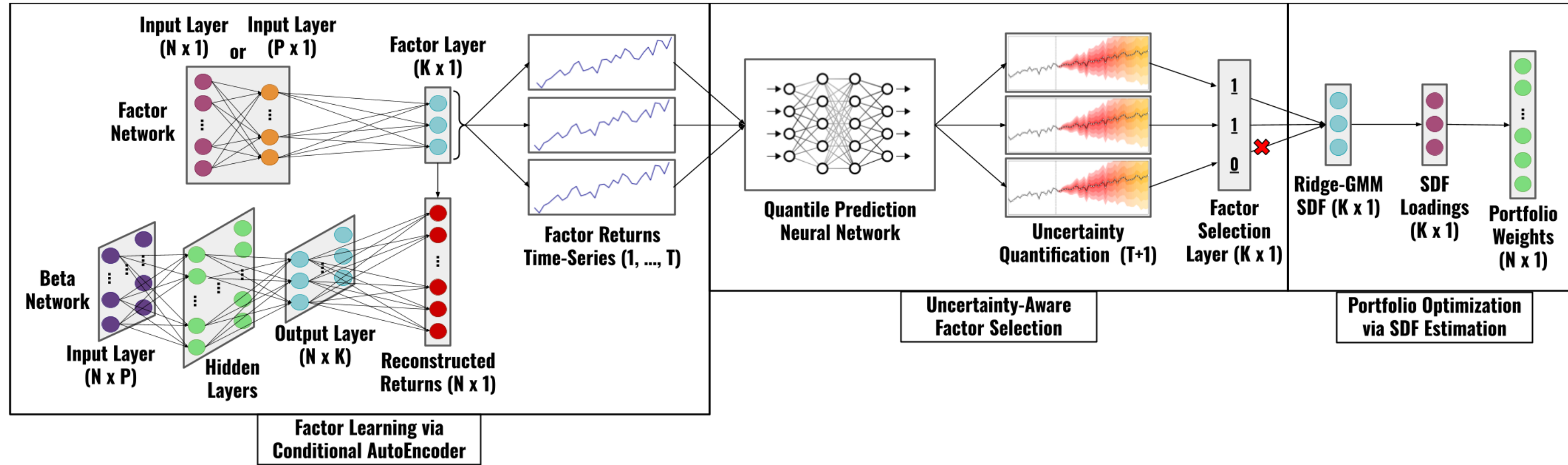
FUSE: FACTOR LEARNING WITH UNCERTAINTY-AWARE SELECTION AND ESTIMATION

Ryan M. Engel, Pawel Polak

Stony Brook University



Introduction



An investable factor model must construct factors, decide which are predictably rewarded, and combine them into a portfolio. Standard pipelines do this in sequence, so each stage optimizes an objective misaligned with the next. FUSE learns all three under one differentiable objective. We make the following contributions:

- **End-to-end learning for asset pricing:** factors, uncertainty-aware selection, and SDF portfolio weights are learned jointly by differentiating through the no-arbitrage pricing problem.
- **A modular uncertainty-aware forecasting interface:** only conditional factor-return quantiles are required, so various forecasters fit the same framework.
- **Decision-focused adaptation of a time-series foundation model:** Chronos-2 is fine-tuned with LoRA, improving risk-adjusted performance over zero-shot Chronos-2.
- **Robust out-of-sample evidence:** on U.S. equities from 2000–2024, FUSE is competitive net of liquidity-scaled transaction costs and produces returns complementary to standard factor models.

Methodology

1. Factor construction. A conditional autoencoder reconstructs cross-sectionally ranked returns with nonlinear loadings $\beta_i(\cdot)$ and latent factors f_s :

$$\text{rank}_{[-1,1]}(r_{i,s}) = \beta_i(z_{i,s-1})^\top f_s + u_{i,s}$$

Ridge characteristic portfolios feed a learnable linear factor map:

$$x_s = \Omega(Z_{s-1}) \text{rank}_{[-1,1]}(r_s), \quad f_s = W_f x_s$$

2. Uncertainty-aware gate. Forecast quantiles at $\tau \in \{0.1, 0.5, 0.9\}$ give a median $\mu_{k,t}$ and a scale $\sigma_{k,t}$, scored as $s_{k,t} = |\mu_{k,t}|/(\sigma_{k,t} + \varepsilon)$:

$$\text{Forward: } g_{k,t}^{\text{hard}} = \mathbf{1}\{s_{k,t} > \vartheta\}$$

$$\text{Backward: } g_{k,t}^{\text{soft}} = \sigma(c(s_{k,t} - \vartheta))$$

Hard zeros cannot be undone downstream, so the gate selects rather than reweights.

3. Differentiable SDF layer. A linear SDF $m_{s+1} = 1 - b_t^\top f_{s+1}^{(\kappa)}$ is priced against $M = 153$ quintile-sorted test portfolios. Its loadings solve a ridge-penalized GMM problem in closed form:

$$b_t = (\hat{C}_{RF}^\top \hat{C}_{RF} + \lambda I_K)^{-1} \hat{C}_{RF}^\top \hat{\mu}_R$$

and map back to stock space through the same projection:

$$w_{r,t} = \Omega(Z_t)^\top W_f^{(\kappa)\top} w_{f,t}^{\text{SDF}}$$

4. Joint objective. All stages train under a single loss, evaluated at the closed-form optimum b_t^* :

$$\min_{\theta, \alpha} \sum_t \mathcal{L}_{\text{price}}^* + \eta \mathcal{L}_{\text{CAE}} + \phi \mathcal{L}_{\text{pinball}} + \rho \mathcal{L}_{\text{gate}}$$

Quantile gradients push the representation toward factors that are forecastable in the first place.

Experimental Setup

- ~4,000 U.S. equities/month; 153 characteristics (Jensen et al. 2023), 1962–2024.
- Train 1962–1999; out-of-sample 2000–2024; monthly rebalancing, retrained every 12 months on an expanding window.
- 72-config grid selected by net validation Sharpe; 50-seed expert ensemble per configuration; 300-stock portfolios.
- ADV-scaled costs: 5 bps for the most liquid names up to 100 bps at the floor. All results are net of transaction costs.
- 8 forecasters: IID-BS, Q-MLP, Q-RNN, Q-GRU, Q-LSTM, Q-TF, Chronos-2, Chronos-2 (LoRA)
- Baselines: SPY, IPCA and CAE at $K=5$ (no selection), 4 staged SCAE variants

Results & Conclusion

Model	Sortino	Sharpe	Ret. (%)	Vol. (%)	DD (%)	Beta	TO
Ensemble (A)	2.39	1.54	14.11	9.14	13.17	0.104	2.67
Ensemble (B)	2.72	1.56	14.59	9.37	20.08	0.053	3.02
IPCA $K=5$	1.88	1.01	9.12	9.04	18.71	-0.087	2.87
CAE $K=5$	2.26	1.28	12.03	9.42	17.49	-0.087	3.21
SCAE (IID-BS)	2.28	1.36	11.21	8.27	17.15	-0.034	3.04
SCAE (Q-Boost)	2.44	1.19	10.31	8.70	13.11	-0.112	3.18
SCAE (Chr-2) [†]	2.24	1.46	13.80	9.43	13.09	-0.088	2.93
SCAE (Chr-2+LoRA) [†]	1.68	1.04	7.96	7.69	20.88	0.130	2.93
FUSE (IID-BS) [†]	2.27	1.37	11.73	8.57	11.94	-0.057	2.25
FUSE (Q-MLP) [†]	1.36	0.85	8.26	9.75	17.40	-0.097	2.44
FUSE (Q-RNN) [†]	1.81	1.02	10.59	10.43	20.93	-0.018	2.71
FUSE (Q-GRU) [†]	2.07	1.00	10.40	10.40	15.29	-0.036	2.33
FUSE (Q-LSTM) [†]	2.15	1.04	9.68	9.30	14.39	-0.020	2.27
FUSE (Q-TF) [†]	2.31	1.13	11.12	9.89	10.43	-0.117	2.38
FUSE (Chr-2) [†]	1.90	1.20	11.23	9.38	17.94	-0.105	2.40
FUSE (Chr-2+LoRA) [†]	2.56	1.31	11.13	8.48	17.01	-0.035	2.25
SPY	0.77	0.55	8.23	15.08	50.79	1.000	—

- Factor construction, predictability, selection, and portfolio formation become one differentiable problem.
- Strongest variants match or beat staged baselines on Sortino and drawdown, and performance holds across very different forecasting architectures.
- LoRA fine-tuning lifts FUSE's Sortino from 1.90 to 2.56, while its lowered it in the staged pipeline.
- FUSE returns are unexplained by standard factors and combine into an ensemble at 2.72 net Sortino.
- IPCA and CAE correlate at 0.65 with each other, but only 0.25–0.45 with FUSE: the objective determines what the representation recovers.

